

NOUN PHRASES AND INFORMATION FLOW IN ESL NARRATIVES: A CASE STUDY

Barbara Adelman

International Institute of Metro St. Louis

Within the field of discourse analysis, cross-linguistic research may lead to broad generalizations about language patterns. Research into second language discourse can support these language patterns and confirm language constraints both for the target language and across language types.

In the attempt to describe the systems underlying real language use, discourse analysts face the issue of how the information density of language is structured to facilitate comprehension. Discourse studies concerned with information status make a crucial distinction between given and new material (Brown and Yule, 1983). Investigation of information flow has shown that given/new configurations are reflected in the grammars of natural languages with individuals varying in their treatments of the given/new categories. (Comrie, 1989). As second language learners struggle to manipulate forms in order to communicate, they reveal varying command of information status in the target language and uncover the strength of constraints operating within the domain of information flow.

Examination of information flow at the clause level in oral narratives has been scrutinized carefully in Prince (1981) and DuBois (1987). They arrive at common conclusions regarding the information status of noun phrases (NP) in the clause core, in particular that the majority of noun phrases convey given information. They also agree that when new information is presented, it is usually in the object position. The information capacity of subject position, however, stands as an area of disagreement. While focusing on information status—new or given—and its effects on the grammatical shape of text, Prince finds new information limited to object noun phrases. DuBois, concentrating on the relationship between information status and syntactic roles within the core of clause, finds that new information occurs not only in subject noun phrases but also has a specific function.

The study by DuBois showed an asymmetrical pattern of placement of lexical noun phrases in oral narratives of an ergative-absolutive language, Sacapultec Maya. These languages mark the Object of transitive verb and the Subject of intransitive verb in the same manner in contrast to nominative-accusative languages which pair Subject of transitive verb and Subject of intransitive verb in opposition to Object of transitive verb. In his study, DuBois concluded that information flow is organized along a Preferred Argument (PAS) structure where lexical NPs and new information lexical NPs are asymmetrically placed across grammatical relations. Native speakers prefer putting new information in the Object position but occasionally use new information in Subject

position. As such the use has a special function—to introduce new characters to the narrative. DuBois, with transitivity distinctions in mind, found avoidance of new NPs in transitive Agent role but not in intransitive Subject. He also articulated other specific constraints involving all lexical NPs, not just those indicating new information : 1) avoid lexical NP in transitive subject position, 2) avoid more than one lexical NP in core clause positions and 3) avoid more than one new lexical NP in core clause positions.

Prince uncovers a similar asymmetrical pattern in a nominative-accusative language study. While not making a distinction between Subject and Agent, she found in a study of English oral narratives that there was a definite tendency to avoid putting new information in subject position. Object position was the vehicle for introducing new information. She further refined the concept of given/new into a continuum in which a speaker is constrained to use the highest level of old information based on his assumption of the listener's awareness. To do other wise is to sound devious or arrogant. She called her model an Assumed Familiarity Taxonomy(AFT). It categorizes NPs into three main types: Evoked, Inferred and New. New is further divided into Brand New, anchored and unanchored, and Unused. Evoked also has two parts, as given information can either be evoked by a situation or by previous text.

In her analysis of informal English conversational discourse, Prince found information status correlated with grammatical relations. The imbalance in formation flow shows overall, subject position NPs were nearly all Evoked, none were New, and 1/15 were Inferred. In Object position less than half were Evoked, 1/6 were New and 1/3 were Inferred.

The combined constraints lead to several predictions concerning naturally occurring discourse produced by second language learners. Specifically, such narrative should contain mostly given information across NPs in the core grammatical roles, Agent, Object, and Subject, consistent with the Assumed Familiarity Scale of Prince and Preferred Argument Structure of DuBois. They should also rely on new information NPs in Object position, leaving subject and Agent to convey largely given information. Given the Preferred Argument Structure and its relation to the lexical form, one would predict that subject should contain some new information lexical NPs

On the other hand, if the constraints DuBois and Prince postulate are not working, the use of lexical NPs and new information NPs would be evenly distributed across all roles in the core clause. It was to test these patterns and constraints under the conditions of second language learning that the present study was formulated.

Method

As set of written English as a second language narratives was collected and analyzed. As in the DuBois and Prince studies, this analysis tracked the use and shape of NPs relative to their potential to carry new information and their ability to

maintain continuous referential identity over time, i.e. given information. NPs analyzed were restricted to the grammatical roles of the core clause: Agent; Object; Subject. This limited syntactic frame offered a clear and measurable field for describing patterns of ESL information flow.

The subject writing the samples was an adult Japanese woman (K), trained as a pharmacist, who had come with her husband and two children to spend two years in the United States. She had had four years of English training in Japan. Although the corpus contained a wide variety of writings, the study was restricted to the narrative discourse type as it was a universal type and was closer to oral story telling styles of the previous studies. Within the narratives, groupings were made according to stimulus type: five with topics prompted by pictures and five prompted by an activity or reading.

Each sample was written, with me, the teacher, as the audience within a 30-45 minute period after a brief class discussion of the prompt. The writing samples analyzed for the study were first drafts, where uncorrected forms could be identified. While the resulting narratives are not transcriptions of oral narratives as in the studies of DuBois and Prince, they do share with the latter many of the same characteristics: informality and certain level of spontaneity, linear temporal development, know and familiar audience, shared background information, and a desire to communicate.

Procedure:

A core clause in the data, consistent with DuBois, was identified as a finite verb having an explicit or implicit (zero) subject. What followed the core clause was considered part of that clause as long as there was no other conjugated verb. This allowed for the identification of lexical and pronominal NPs by quantity and position according to the roles Agent, Object and Subject. To isolate Agent from Subject, clauses were identified according to a transitive/intransitive dichotomy. By using surface level relations, a verb was classified as intransitive if it accepted any constituent other than a NP. A verb followed by an infinitive or participial complement, a sentential complement, or prepositions was considered intransitive. The verb to be was categorized separately.

After core NPs were established and organized according to quantity and position, their information level was assessed. The pragmatic indices, New, Given and Inferred as defined in Prince's Assumed Familiarity Taxonomy were used to classify core NPs in the ESL data. The classification was done twice, three months apart. The second analysis confirmed the classifications for NPs in the Non-Picture Prompt (NPP) narratives. However, there was some variation categorization of NPs in the Picture Prompt (PP) narratives

What was graphically shown in the picture was considered Situationally Evoked (SE) unless it was referred to again, in which case it was Textually Evoked (TE). New information was considered anything the writer added to the picture. Therefore, articulating relationships between people in the picture, giving them names and a past history were all considered Brand New (BN).

New information NPs were tracked as to quantity and position in order to test whether K avoided more than one new argument per clause and whether she avoided a new referent in Agent position. Once a pattern of information flow was established, the noun phrase form was more closely examined.

The syntactic shape of NPs, new NPs in particular, was recorded. Lexical NPs which can be preceded by a variety of markers were noted in conjunction with their marker. As the NPs had previously been categorized according to AFT, the new lexical NPs and their marker could easily be identified. All NP markers were tabulated in order to tie the classification of new information to the grammatical domain, thus testing the hypotheses of Prince and DuBois in relation to the asymmetry of information flow.

An error analysis of new lexical NPs and the markers they contained was made. The resulting data was used to test the validity of the Familiarity Scale (FS) construct as applied to K's writing sample.

Finally, the narratives were surveyed for syntactic constructions which shifted new information out of Subject and Agent roles, since Prince asserts that the native speaker employs certain syntactic construction to maintain an imbalance in information flow. Once isolated, these constructions were classified and compared to native speaker constructions Prince had identified in her study. They were also examined relative to their impact and function in the wider discourse in order to test DuBois hypothesis that native speakers assign a special introduction function to the Subject role.

Results

Quantity and Frequency of Lexical Noun Phrase Use

Overall frequency counts of noun phrase shape in the writing samples reveal constraints on the number and placement of lexical NPs within the clause core. The data contain 237 clauses; therefore, the possible occurrence of lexical NPs could range from 0 to 474 (if all clauses were transitive). However pronoun NPs, 220, are frequently used in the core; while only 100 lexical NPs are used. Even if all lexical NPs were new, there would still be more given information in the narratives which supports one of the patterns found by Prince and DuBois.

The possible distribution of NPs within any one core clause could result in zero, one, or two lexical NPs per clause. The pattern which results from the count of location

of NPs. Across both stimuli, 92 clauses contain lexical NPs representing 37% of the total 237. Of those 92, only four contain a lexical NP in both grammatical positions, which is 1.7% of all clauses and 5% of clauses containing lexical NPs. This establishes the rarity of two NPs in a clause. Further, while clauses hardly ever exceed more than one lexical representation in a core NP position, the number of clauses with zero lexical NPs (145) is almost twice the number containing one lexical NP (88).

A lack of transitive constructions could bias the quantity count, since that would limit the possible appearance of two lexical NPs per clauses. Across stimuli type samples, there are 89 transitive and 148 intransitive clauses. That is, 37% of the corpus is transitive. This creates ample opportunity for two lexical NPs to appear together in the core. All else being equal, the relative low representation of two core lexical NPs, 1.7%, reflects a skewed distribution between the two clause types. This supports the tendency to suppress lexical NPs in core argument positions.

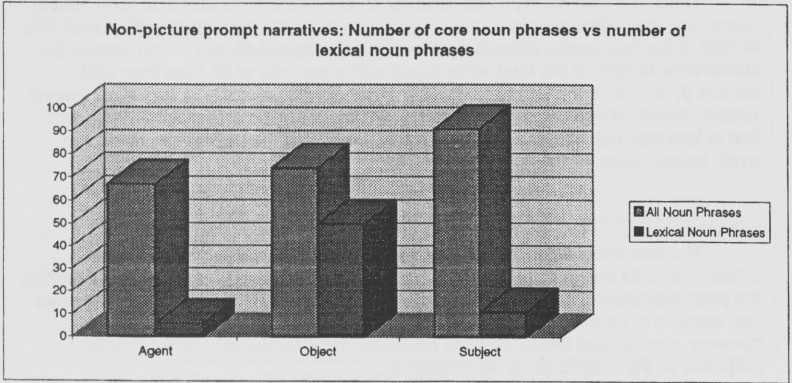
If transitive/intransitive clauses still support an asymmetry in the distribution of lexical NPs, perhaps another construction could have an effect on the quantity count: the predicate nominative, where two lexical NPs have the possibility of appearing in conjunction with a finite verb. However, in the corpus only thirteen examples of the construction exist and only one of those uses lexical NPs in Subject and predicate nominative slots. These data also support the overall preference shown previously for avoiding more than one lexical argument in a clause. The ESL data show a preference for less than two lexical NPs in core argument positions 97% of the time and less than one 62.2% of the time. The quantitative data clearly demonstrate the existence of a constraint on the number of lexical NPs in the clause core which supports the native language constraints found by Prince and DuBois.

Preferred Placement of Lexical and Pronominal Noun Phrases

Assessment of the role or roles that lexical NPs entertain within the core shows that lexical NP use is similarly non-random. The relative capacity of Agent, Object, and Subject compared to the actual distribution of lexical NPs uncovers the constraints on placement. The distribution of lexical NPs in the writing samples shows that a major proportion are in the Object role.

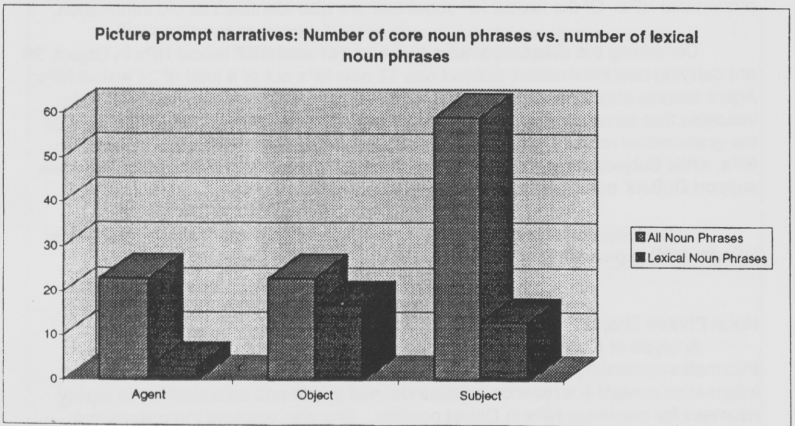
The data are arranged according to stimulus prompt in graphs 1 and 2 in a syntactic profile of the narratives. The graphs show the total number of core noun phrases versus the number of lexical noun phrases by syntactic position.

GRAPH 1



| | All Noun Phrases | Lexical Noun Phrases |
|---------|------------------|----------------------|
| Agent | 67 | 6 |
| Object | 74 | 50 |
| Subject | 91 | 11 |

GRAPH 2



| | All Noun Phrases | Lexical Noun Phrases |
|---------|------------------|----------------------|
| Agent | 23 | 3 |
| Object | 23 | 17 |
| Subject | 59 | 13 |

There is a total noun phrase capacity of 332 places filled with pronouns, lexical nouns and zero. Object which has a capacity for 96 (29%) NPs, attracts 67 lexical NPs, or 66% of the total lexical NPs. Subject with a capacity of 146 NPs (44%) attracts 24 lexical NPs, or 23% of the total; while Agent with a capacity of 90 (less than 25%) attracts 9, or 10% of the lexical NPs. The distribution of lexical NPs is skewed toward Object. Agent, whose capacity over half of Subject's, has an actual lexical NP count that is less than half that of Subject. The ESL data confirms DuBois's constraint to avoid lexical Agent.

Information Flow

The data show a predominance overall of given information realized in pronominal units and in Agent and Subject positions specifically. The ESL data confirm the preponderance of new information in lexical NP form in Object position. The data has samples of new lexical NPs in Subject position and a few in Agent position; however, new Subject lexical NPs are employed more frequently and for specific purposes as the DuBois study would predict.

The incidence of pronoun use, assigned given status by the Assumed Familiarity Taxonomy, in all of the Agent positions is 82% (n=57) in NPP and 81.8% (n=18) in PP. In Subject position, it is somewhat lower: 81% (n=73) in NPP and 72.8% (n=43) in PP. Object however, employs pronouns in possible slots only 32% (n=23) of the time in NPP and 26% (n=6) in PP. This assessment of given status includes only the pronominal NPs. Of the lexical NP about half are also identified as old information.

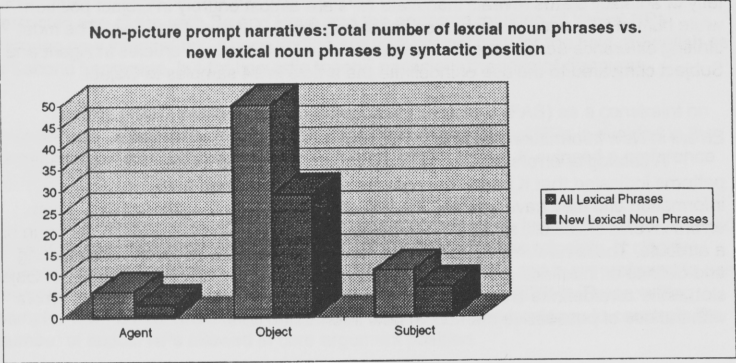
Combining the data for stimulus types, out of a total 67 lexical NPs in Object, 36 are carrying new information; Subject has 12 new NPs out of a total of 24 lexical NPs; Agent attracts only 5 new NPs out of a total of 9 lexical NPs. Again the asymmetry indicates that some constraints are operating on the attraction of new lexical NPs by the grammatical roles. Object by far attracts the largest number of new information NPs, while Subject attracts far less, it does attract twice as many as Agent. The data support DuBois' constraint to avoid new lexical Agent.

A summary the tabulation of NPs carrying new information versus total lexical NPs for NPP is given in graph 3 and for PP in graph 4 on page 8.

Noun Phrase Shape

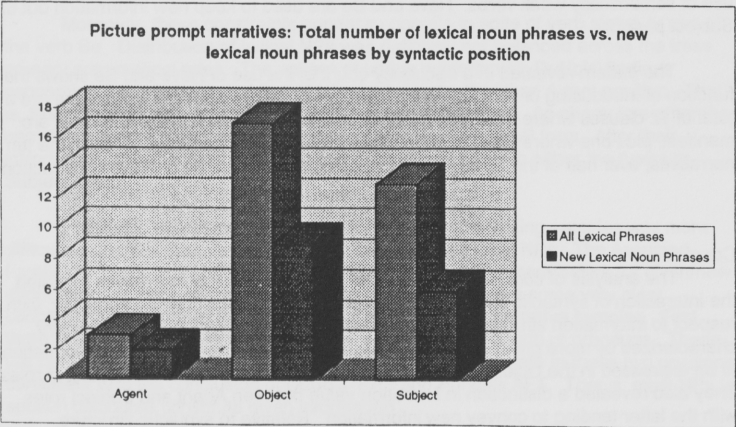
Analysis of K's writing samples demonstrate a preference for marking new information according to grammatical role. A profile of NP shapes expressing new information reveals K's reliance on the indefinite article and zero marking to signify newness for inanimate NPs in Object position. She also employs the possessive pronoun across roles to create new NPs.

GRAPH 3



| | All Lexical Phrases | New Lexical Noun Phrases |
|---------|---------------------|--------------------------|
| Agent | 6 | 3 |
| Object | 50 | 29 |
| Subject | 11 | 7 |

GRAPH 4



| | All Lexical Phrases | New Lexical Noun Phrases |
|---------|---------------------|--------------------------|
| Agent | 3 | 2 |
| Object | 17 | 9 |
| Subject | 13 | 6 |

A count shows zero or possessive markers to be the most frequently used. A tally of animacy status reveals inanimate NPs are almost entirely in Object position while human NPs occur across roles, most frequently in Subject position. The most striking difference across roles is the complete lack of indefinite articles in Agent and Subject compared to the use of indefinite the article in 14 samples in Object.

Errors in New Information Marking

A catalog of errors marking new information NPs was made. The resulting patterns indicated that K based her choices on core role when identifying new information and may have equated the adjective and article in marking NPs. Most errors involve omission, six of ten. However, omission of "a" occurs when the noun has a attribute. These new information errors focus on the choice between zero marking and definite or indefinite articles. The zero/indefinite choice creates problems in Object slot, while zero/definite creates problems in Agent and Subject. There are no errors with the use of possessive markers in new information NPs.

Preferred Syntactic Constructions

Statistically, K's data support a pattern of information flow with a concentration of new information in the Object role. However, the Agent and Subject roles, while informationally weak, do provide some new information. Because K does not have in her grammatical repertoire constructions like it-clefts or left dislocations and uses the presentative there construction only twice, this imbalance in roles is sustained by the use of semantically weak verbs. Have and Be are used to keep new information out of Subject position

The pattern revealed in a frequency count of the use of Have and Be shows the function of introducing or concluding the narratives. Across stimulus types there are a total of 12 clauses where the introductory or concluding function of Be and Have are manifest, plus one where Have changes the topic within the narrative. Thus in the ten narratives, over half of the 20 beginnings and endings involve Be or Have constructions

Conclusions

The analysis of core noun phrases (NP) in ten written ESL narratives revealed the interaction of syntactic shape and position with information flow. In particular, with respect to information structure, the narratives showed information flow asymmetry, characterized by more given items than new, and by distinct tendencies for information to be expressed in the Object role and given information in the Agent and Subject roles. They also revealed a distinction in formation value between Agent and Subject roles with the latter tending to convey new information. Relative to syntactic structure, narrative analysis uncovered the limited use of lexical NPs in the clause core, noting in

particular the avoidance of lexical NPs in Agent position. NP quantity and use to prompt Lastly, the narrative highlighted maintenance of information flow asymmetry though restricted use of the verb Be and Have and the preferred marking of lexical NPs carrying new information. These findings are relevant not only for the field of English as a Second Language, but for linguistic theory, especially theories of discourse.

DuBois (1987) posited Preferred Argument Structure (PAS) as a constraint on core NP use. Essentially this principle states that use of lexical NPs is avoided in the Agent role and that constraints operate in four specific ways. K evinced a preference for a principle similar to PAS in her narratives.

The essential feature of PAS is a number constraint on lexical NPs, that is, avoid more than one lexical NP per clause core. Consistent with this principle, K clearly preferred to use fewer than two lexical NPs per clause (98%) and even preferred to use fewer than one lexical NP per clause more than half the time (62%). These writing samples from an intermediate ESL speaker thus confirm the tendency to constrain the number of lexical NPs allowed in core argument position.

The PAS constraint on placement of lexical NPs, i. e. avoidance of lexical NP in Agent role, is also supported. The Agent role accounts for only 10% of lexical NPs, while Subject position support over twice that amount (23%) and Object, 66%. The ergative-absolutive alignment of Subject with Object as opposed to the nominative-accusative alignment of Agent with Subject is not as strong in these written ESL narratives as in DuBois' Sacapultec Maya but the tendency definitely appears.

Moreover, these constraints appear to operate in spite of verb status or use of the verb Be. Distribution of lexical NPs was definitely not balanced across the three primary grammatical roles. The pattern of usage conform with DuBois' findings concerning the greater use of lexical NPs in Subject rather than Agent position. That is, Subject works as a staging platform for actors in narrative. As characters enter the narrative they are introduced in the Subject role in a lexical NP form. After their introduction, they become given information and thus occur in either the Agent or Subject roles.

Pragmatic assessment of information status in the writing samples was more difficult to quantify than the number and placement of lexical NPs. The Assumed Familiarity Taxonomy (AFT) of Prince (1981) helped set the parameters for given/new information. In spite of some categorizing difficulty, there are some very clear conclusions to be drawn from a pragmatic analysis of the narrative samples. First, there is an overriding use of Evoked NPs, especially in the form of pronouns. These are not limited to Agent and Subject roles, but they do dominate in these two positions, consistent with the conclusion of DuBois' PAS and Prince's AFT. That is, most NPs in spoken discourse constitute given information.

Overall, the data generally confirm the constraints noted in AFT and PAS: a speaker avoids more than one new NP per clause and avoids placing it in Agent position. In the data from K, clearly the most restricted use of the new NPs is with Agent at 9% (n=5), followed by Subject, 23% (n=13), and Object, 68% (n=39). Subject is used over twice as much as Agent for new NPs. There are, however, a few cases of new lexical NPs in non-Object position that neither Prince nor DuBois found. Still, one can conclude that the PAS and AFT constraints operate fully in these ESL data, although perhaps with a somewhat diminished intensity.

The patterns emerging from NP marking demonstrate a preference according to grammatical role and reasoned choices in marking information flow. In general, error analysis indicates a secure concept of the indefinite article to introduce new NPs in Object slot. K tends to undermark the NPs. The construction adjective + noun especially presents problems for K, both by omitting the definite and indefinite article and exchanging "a" for "the." This could be first language interference.

The position of new anchored and unanchored NPs appears to be fairly well established in K's usage. For both Agent and Subject roles, K preferred Brand New anchored NPs. For Object, however, she preferred unanchored new NPs. Most of the anchored NPs are marked with possessive pronominal modifiers. As none of the new NPs marked this way had no incidence of error, this strategy was a successful means for introducing a new NP. She avoids choosing between the articles where chances of error exist.

Whatever K's difficulties with marking the information status of noun phrases, she does have a sense of the broader principle of narrative discourse construction. The need to frame a topic was apparent in the data as evidenced by her use of stative verbs in introductory and concluding sentences. She demonstrates an awareness of cohesion in the use of pronouns. But these are about the only strategies she uses. This emphasizes the conflict that often arises in second language learning: the rich knowledge and information available for communication and the impoverished level of skills with which to communicate.

REFERENCES

- Brown, Gillian and George Yule. 1983. Discourse Analysis. Cambridge: Cambridge University Press.
- Comrie, Bernard. 1989. Language Universals and Linguistic Typology. Second edition. Oxford: Basil Blackwell.
- DuBois, John W. 1987. The discourse basis of ergativity. Language, 63(4), 805-855.
- Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (Ed.) Radical Pragmatics. New York: Academic Press. (223-255).